Party Facts: A database of political parties worldwide

Döring, Holger; Regel, Sven

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Döring, H., & Regel, S. (2019). Party Facts: A database of political parties worldwide. *Party Politics*, *25*(2), 97-109. https://doi.org/10.1177/1354068818820671

Döring, Holger; Regel, Sven

Article — Published Version

Party Facts: A database of political parties worldwide

Party Politics

Provided in Cooperation with:
WZB Berlin Social Science Center

Suggested Citation: Döring, Holger; Regel, Sven (2019) : Party Facts: A database of political parties worldwide, Party Politics, ISSN 1460-3683, Sage, Thousand Oaks, Vol. 25, Iss. 2, pp. 97-109,
http://dx.doi.org/10.1177/1354068818820671

This Version is available at:
http://hdl.handle.net/10419/209736

Article

Party Politics

Party Politics
2019, Vol. 25(2) 97–109

DOI: 10.1177/1354068818820671
journals.sagepub.com/home/ppq
SAGE

Party Facts: A database of political parties worldwide

Holger Döring
University of Bremen, Germany

Sven Regel
WZB Berlin Social Science Center, Germany

Abstract

Here, we present Party Facts (www.partyfacts.org), a modern online database of political parties worldwide. With this project, we provide a comprehensive database of political parties across time and world regions, link party information from some of the core social science data sets, and offer a platform to link political parties across data sets. An initial list of 4000 core parties in 212 countries is mainly based on four major data sets. The core parties in Party Facts are linked with party information from some of the key social science data sets, currently 26. From these data sets, we have included and linked about 15,000 party observations. Party Facts is an important step in developing a more coherent operationalization of political parties across time and space and a gateway to existing data sets on political parties. It allows answering innovative party research questions that require the combination of multiple data sets.

Keywords

collaborative data collection, data sets, political parties

Introduction

Political parties are a core unit of analysis in social science research and a large amount of information about them has been collected. Definitions and sampling strategies for political parties as units of analysis vary significantly, party names are recorded in very different ways (e.g. abbreviation, original language name, English name, ambiguous spelling), and the discipline is lacking common global party identifiers. The vast amount of information about political parties is also difficult to combine and a significant amount of time is spent on harmonizing and combining party information from different sources. However, innovations in party research increasingly require the combination of multiple party data sets, a mundane, time-consuming, and error prone part of empirical work.

Challenges in defining, measuring, and harmonizing data on political parties are in line with a more general debate on the dilemmas and failures in cross-national data collection. Schedler (2012) has emphasized that cross-national data collection in social science research often relies on opaque concepts, varying sources and different coding procedures. He also points out that data duplication and data set incompatibilities are major challenges in cross-national research. Research on political parties represents a prominent example. Researchers should use modern approaches to harmonize data in party research and need to rely on recent technological innovation (Döring, 2013; Mustillo and Springer, 2014). Harmonized party information is crucial for a variety of research questions that rely on different data sets with party-level information. It is also a prerequisite for addressing current research questions that are in need of multiple sources which can easily be combined.

We have developed Party Facts, a modern database of political parties across the globe and present a new approach to data collection. Party Facts is a modern

Paper submitted 4 April 2018; accepted for publication 27 November 2018

Corresponding author:
Holger Döring, SOCIUM Research Center on Inequality and Social Policy, University of Bremen, Mary-Somerville-Straße 5, Bremen 28334, Germany.
Email: doering@uni-bremen.de

successor to classical party almanacs. It provides a set of core information about relevant political parties worldwide and links parties across existing data sets. Party Facts already includes many of the core social science data sets with party-level information. With Party Facts, we make use of recent technological innovations in data storage and Web development to provide an easily accessible database of political parties. A database allows us to structure the data systematically, and a modern Web interface enables registered users to link parties across data sets and to revise existing information about the parties included in Party Facts. A long-term data archive provides independent and permanent access to the party data. Currently, Party Facts covers around 4000 parties and links about 15,000 parties from more than 25 external data sets. The initial set of core parties is mainly based on information from four data sets: ParlGov, Manifesto Project, CLEA, and PolCon. These links can be used to merge data sets and to use relevant indicators contained in these external data sets for innovative research questions requiring information from multiple sources.

Here, we present Party Facts in six steps. First, we briefly summarize existing approaches to establishing the population of political parties and present main data sets with information on parties in social science research. Second, we introduce the main characteristics of Party Facts. Third, we discuss the definition and operationalization of political parties in the literature and in Party Facts. Fourth, we present the population of political parties established in Party Facts and its coverage of time and regions in the world. In the fifth part, we briefly discuss the technical implementation of the project and its use of a database, a version control system, and a website. In the sixth part, we present an application of the project and validate different left-right measures. We conclude with a brief summary and suggest future extensions of the project.

Information about political parties

In social science research, we regularly rely on information about political parties and combine different sources. Party-level information may include electoral performance, access to government, public opinion, basis of leader support, or party positions. Political parties, defined as organizations that seek votes in elections to implement policies, are at the core of social science research. However, they are also ambiguous entities. Party information is scattered across multiple sources, these sources are hard to combine, and harmonization of party units is difficult.

There used to be a tradition of systematically collecting information about political parties in handbooks and almanacs. These sources listed all relevant parties for a particular region and time period and provided detailed descriptions about the parties' life cycles. The Political Parties of the World series aimed at providing a comprehensive discussion about all parties across the globe over time (e.g. Alexander, 1982). Others have focused on particular regions, such as Jacobs (1989) for Western Europe and Bugajski (2002) for Central/Eastern Europe, or on a specific country only, for example, Stöss (1986) on Germany. Today, it is only the regularly updated Political Parties of the World (Sager, 2009) that follows this traditional approach. Among social scientists, this type of work on standard references about political parties has been in steady decline or is the by-product of handbooks about electoral results, such as the Nohlen series (e.g. Mackie and Rose, 1991; Nohlen and Stöver, 2010; Rose and Munro, 2003). Nevertheless, these printed sources are often still the main information about political parties within the research community.

Recent decades have seen the emergence of new digital sources with systematic information about political parties. Among them, the Manifesto Project (MRG/CMP/MARPOR) is one of the most prominent sources of digital information about political parties (Budge et al., 2001; Klingemann et al., 2006; Volkens et al., 2013, 2018). It provides a detailed coding of election programs (manifestos) according to a set of defined policy categories. Information from its data can be used to measure the salience of particular issues and to derive party positions. The Manifesto Project has also become a de facto standard for information about political parties in the postwar period. The data from the project have been thoroughly scrutinized over the years as a result of its wide usage within the social science community. The Manifesto Project focuses on parties that gained seats in parliament and does not include the prewar period. Another source, ParlGov, includes all parties that won 1% of the vote in democratic national elections since 1900 (Döring, 2016). Nevertheless, the Manifesto Project and ParlGov focus on (semi-) advanced democracies and a significant number of parties across the globe are not included in these key data sets.

There have been a number of other approaches to measuring party preferences in recent decades. Through this work, we have now a very rich and fine-grained set of information about the positions of political parties. Party expert surveys have been particularly prominent. With this approach, preferences are measured through expert assessments about party positions and issue salience. The first party expert surveys focused on the left-right dimension only (Castles and Mair, 1984; Huber and Inglehart, 1995; Morgan, 1976), but others quickly extended the approach and included questions about the positions and salience on particular policies (Benoit and Laver, 2006; Rohrschneider and Whitefield, 2012). Today, there exists a set of party expert surveys that focus on broad aspects of party competition (Laver and Hunt, 1992) as well as on more particular aspects (Szöcsik and Zuber, 2015). Among the expert surveys, only the Chapel Hill Expert Survey (CHES) Series is conducted at regular intervals (Bakker et al., 2015). In

applied research, party position data from expert surveys are regularly combined with public opinion surveys but there is little systematic effort to harmonize the many existing sources.

Public opinion surveys include information about parties in vote intention questions and in questions about the party chosen in the last election. Some surveys, such as the European Election Studies (EES), also include a "propensity to vote" question and measure a voter's preference toward all main parties in a country. This survey-based information about political parties is a core element of studies of democratic representation. The World Values Survey (WVS) has the broadest global coverage, includes 84 countries, and was first conducted in the early 1980s (World Values Survey Association, 2015). The European Social Survey (ESS) is another prominent source, covering 31 European countries and having its first wave in 2002 (ESS, 2017). Voter preference measures of political parties from public opinion surveys are regularly combined with other sources, such as expert surveys or the Manifesto data. A prominent example of the usage of these combined sources is the congruence debate (c.f. Powell, 2009). Later, we present an application of Party Facts and combine public opinion surveys, expert surveys, and Manifesto data to validate left-right positions.

Other important past and contemporary comparative contributions address party organizational issues in-depth. Janda's (1980) cross-national study, the Data Handbook on Party Organizations by Katz and Mair (1992), and the more recent Political Parties Database PPDB (Poguntke et al., 2016) provide information on a wide range of party organizational characteristics. There is also important information about political parties in subfields of comparative politics beyond pure party research. For example, Andersson et al. (2014) have collected information on coalition governments in postwar Western Europe with detailed information about political parties at each instance of government formation. Information about political parties is only a by-product of these sources on democratic politics but nevertheless constitutes systematically collected data on parties. For research on political parties, this wider set of sources is invaluable to answer specific questions in party research, but it is often difficult to integrate these sources with the mainstream data sets presented above.

The rich set of existing information on political parties we have presented is most valuable if it can be easily combined across data sets. Applied work is often dependent on combining party-level information about voters, politicians, and public policies. Drawing on existing sources allows us to establish the population of relevant political parties in the world across time. However, existing sources constitute something better described as a Tower of Babel, with different definitions of parties and distinct unique identifiers used across data sets. It is tedious or simply frustrating to try to combine existing data sets on political parties. Hence, a fresh approach to combining and systematizing existing sources is an important contribution to social science research.

Party Facts: A database on political parties

With Party Facts, we present a new database on political parties in the world that combines party information from important social science data sets and aims to solve the collective linking dilemma. In Party Facts, we currently include around 4000 core parties in 212 countries, mainly based on four major data sets: ParlGov (Döring and Manow, 2018), CLEA (Kollman et al., 2018), PolCon (Henisz, 2000), and Manifesto Project (Volkens et al., 2018). The coverage of parties in the different world regions varies, and Figure 1 gives a summary of the world coverage in Party Facts. Each country is colored according to the number of covered parties. Figure 1 shows that we cover all world regions and include many parties for the countries in Europe and the Americas but fewer parties for African and Asian countries.

In the project, we distinguish between core parties and external parties. Core parties are the party units newly created and edited in Party Facts. They are our approximation of the population of all relevant political parties of the world and a starting point for an authoritative reference list. External parties are the party observations extracted and imported from data sets and subsequently linked to our core parties. The core parties are fully linked to about 15,000 external parties in more than 25 social science data sets. Among others, these external data sets include expert surveys, party organization data, party content analyses, public opinion data, and election results.

The data sets included in this initial presentation of Party Facts have been linked by trained coders using the Party Facts codebook and the advanced party-linking tools of the Party Facts website. Most of the linking is based on the party names only. The coders used additional information (e.g. party strength) from the data sets if a party name was not sufficient to link an external party. All party links can also be validated on the website by registered users.

Party Facts relies on a technical approach that combines a database, a version control system, a website, and long-term archiving. This new technical approach and the tools are described in detail in one of the later parts of the article. In Party Facts, we store all party information about core parties and external parties coherently in a database. Party information from data sets is extracted with scripts. These scripts and the party information we import are stored in a version control system. A version control system allows us to document and archive different versions of the data we import into the database. The website provides a platform for collaborative online editing of the party information in Party Facts and gives access to the recorded data. All information and data on core and external parties provided by

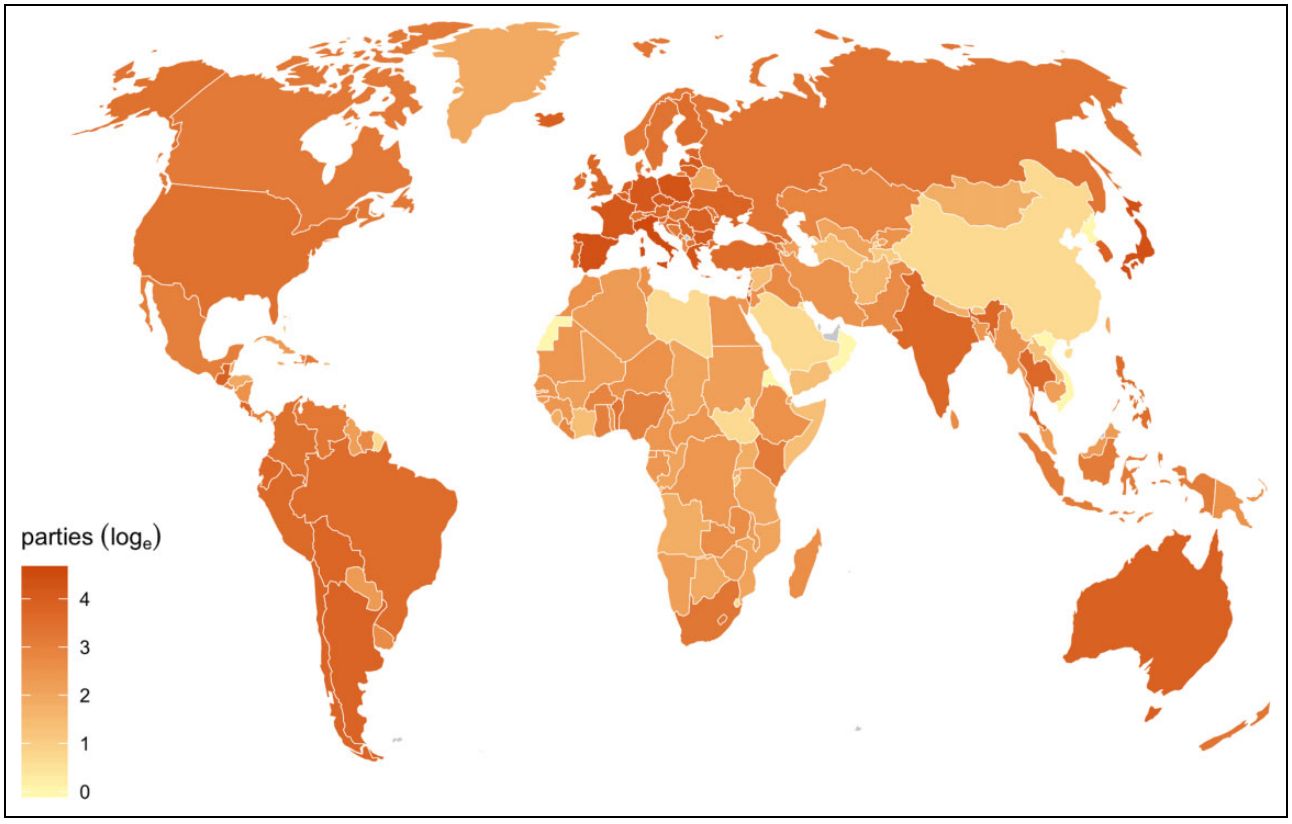

Figure 1. World map of Party Facts core parties. *Note*: Number of core parties (logarithm) in countries covered by Party Facts.

Party Facts include a time stamp about the most recent modification to track changes. The data we collect with Party Facts are also stored in a long-term data archive at regular intervals. Our novel approach to combine party information from different sources increases the transparency of data linking and fosters collaborative data collection.

Party Facts is an important step toward developing a comprehensive record of political parties in the world. The project provides an important supplement and a modern extension to almanacs on political parties. It supports harmonizing the many existing data sources about political parties. In the long run, Party Facts permits the development of more coherent coding rules for political parties as units of analysis, by establishing the population of parties, as it is recorded in major social science data sets. The next sections present our harmonization strategy as well as the population of parties we have collected.

Defining and measuring political parties

In Party Facts, we link and harmonize existing party information from different data sets. This linking approach allows us to establish the population of relevant political parties in the world and across time. Such an approach should ideally rely on a coherent operationalization of political parties as units of analysis. Parties often change their names several times over their history. A political party may also run under a new name, as part of an electoral alliance, or may form only temporarily during a parliamentary term. However, existing sources define and record these changes of political parties differently.

Even more surprisingly, there is no established operationalization of a new political party within the research community that allows us to define and code new parties coherently. Many new parties are formed from existing parties. They may only change their name, or a merger may be dominated by a large party. Barnea and Rahat (2011) provide a detailed discussion about several challenges in defining and measuring new parties. They highlight the fact that party change can occur in different dimensions: changes can occur in ideology, the organization, the electorate, or among activists or the elite. They propose a multidimensional framework to assess and measure the "newness" of parties and provide a dichotomous measure of new parties. It has yet to be seen whether these definitions can easily be applied to code political parties across different regions of the world and across time. The multidimensional framework proposed by Barnea and Rahat relies on different measures that are potentially hard to identify for newer democracies and across time. Even the dichotomous measure they propose, "no more than half of its top candidates originate from a single former party"

(p. 311), is difficult to establish for all parties. It requires detailed candidate list information which is not easily available for all relevant political parties in the world and time.

We have not solved this puzzle of defining and operationalizing political parties directly. Instead we have created a new list of core parties based on the political parties recorded in existing data sources. Such an approach allows us to establish the universe of relevant political parties as it is currently covered within social science data sets. By creating a new list of core parties, following only some general coding rules, and by linking these parties to existing sources, we have established a starting point for a new discussion about a feasible operationalization and coding of political parties. External data sets often provide additional information about parties, such as electoral results or party positions, which can be used to assess the newness of a particular party.

In Party Facts, we also need a selection criterion for relevant political parties. Most political parties are very small, exist only for a very short period of time, and have little political relevance. Janda (1980: 7) uses a 5% threshold in two successive elections. ParlGov includes election results for all parties that won at least 1% of the vote in national elections (Döring, 2016: 538). This well-defined population of political parties in ParlGov shows that most parties have a maximum vote share below 5% and take part in only one or two national elections. Van de Wardt and Otjes (2018: 9–10) establish the population of all parties that contested elections between 1974 and 2010 in 13 Western European countries and also demonstrate that most new parties are very small. We have decided to apply a general threshold of at least 5% in two national elections to establish the population of relevant political parties. For linking data sets with Party Facts, we use a more flexible threshold. We also include parties that won at least 1% if they are in multiple data sets or if they are available with high-quality party information, as we discuss in detail in the next section.

Defining and operationalizing new political parties and establishing the population of all relevant political parties remain an open field of party research. We are convinced that the pragmatic approach we apply in Party Facts is the only viable starting point for a modern digital almanac on political parties and helps to renew the debate about the definition of new parties within the research community. The newly gathered list of core parties in Party Facts based on existing sources provides the basis for building upon and for improving existing definitions and operationalizations. It also demonstrates the challenge to develop criteria for distinguishing parties that can be applied to all political parties in the world.

Party and data set population in Party Facts

For Party Facts, we have included and linked many of the main data sets in social science research to establish the population of relevant political parties. On the basis of existing data sets, we have created a new list of core parties that covers all countries in the world and over time. We distinguish these newly created core parties in Party Facts from external parties, which are those that we have imported from other data sets. External parties consist of the party information (name, size, duration) that we extract from data sets. Core parties provide the main link between external parties in data sets. Information about the core parties can also be edited in the database. External parties' information is only imported but not altered in the database. It is necessary to distinguish between core parties and external parties, as only creating core parties avoids redundancy in party linking, provides authoritative harmonized party entities with regard to party information such as original name, English name, and life span, and allows the creation of an authoritative unique identifier.

Table 1. Distribution of Party Facts core parties across four world regions.

Continent	Countries	Parties	Age (mean)	Share (mean)
Africa	57	571	15	13
Americas	46	760	15	10
Asia-Pacific	60	959	10	8
Europe	49	1794	10	6

Note: Number of countries and core parties by continent included in Party Facts with mean party age (difference first and last year) and mean maximal vote share by continent.

The depth and global coverage of our core party list varies by region, although we aim for all relevant parties worldwide. This is due to differences in the number of countries that external parties from major data sets cover. There are very few data sets that provide information about the majority of countries in the world or that cover democracies and autocracies as well as long time periods. Most data sets focus on party competition in advanced European democracies. Nevertheless, we have assembled a large and comprehensive set of relevant political parties in the world (core parties) on the basis of existing sources (external parties).

At this stage, Party Facts' core party list includes parties in 212 countries. Table 1 provides information about the regional distribution of these parties in four world regions. About 2000 parties, half of all core parties in Party Facts are from European countries, and the information on the average vote share demonstrates that they include many smaller parties. For Africa, we include about 500 core parties, the smallest number among the four regions. This is due to the fact that democratization in Africa occurred later than in other regions and that there are fewer existing data sets that cover the region. For each of the other two world regions, Asia-Pacific and the Americas, we include about 800 core parties. The number of countries in each of the

four regions does not differ that much and we include about 50 countries for each region.

Our initial records are based on the two major data projects that we have been involved with ParlGov and the Manifesto Project. These two data sets include information about electoral results, and the respective party information has been improved over several years. This information has allowed us to identify parties very precisely. Party information in ParlGov is based on a coherent coding rule of new parties and includes information on predecessor and successor parties (Döring, 2016: 539). The Manifesto Project data cover more countries than ParlGov and have been heavily scrutinized and improved over the years. However, ParlGov and the Manifesto Project have a strong focus on established democracies in Europe and do not cover other regions of the world systematically.

For countries in other world regions, the majority of the information in Party Facts is mainly based on harmonized party information from CLEA and PolCon. The CLEA project has collected district-level results for elections in 156 countries. This information can be used to derive information about the life cycle of political parties and their electoral performance. The PolCon data set combines information about the seat composition of national legislatures (lower and upper chamber) and the partisan affiliation of the head of government and the head of state to derive an index of veto power. It is an extensive set of information containing crucial information about political parties, such as parliamentary strength and access to top executive positions in 181 countries. CLEA and PolCon do not provide national-level vote shares directly, but they still allow us to get some information about the strength of a party over some time period. Party size information from CLEA is calculated on the basis of district-level votes, which may differ slightly from national-level vote shares. The party size information in PolCon is based on a party's seat strength in parliament, and the quality of this information differs across the data set.

These four data sets are the sources for most of the core parties in Party Facts. About half of the parties are originally imported from ParlGov and CLEA. ParlGov only covers 39 countries, so the majority of countries from around the globe have originally been imported from CLEA, which covers 156 countries. In addition, the Manifesto Project and PolCon account for a large share of political parties imported into Party Facts. PolCon has the most extensive coverage of political parties across the globe (181 countries) and accounts for the majority of very small countries (see also Table 2 below). We include different thresholds to filter small parties from these four data sets due to the different detail of party information the sources provide. For ParlGov and Manifesto data we apply a 1% threshold, for CLEA a 2% threshold, and for PolCon we select all parties that won at least 5% of the seats in two elections. Therefore, Party Facts includes all parties in the world that won at least 5% seat share in two national elections and provides a coverage of up to 1% for country data based on ParlGov, Manifesto Project, and CLEA.

In Figure 2, we present the distribution of the political parties from the external sources we have included and fully linked. The graph shows the number of parties for every data set divided into the four major world regions. The number of parties we include to some extent corresponds to the waves and levels of democratization. The earlier a majority of countries in a region transform into democracies, the more external parties we find in Party Facts.

We have included and linked party information from many important social science data sets into Party Facts. Table 2 provides information about these data sets and the number of parties we include. Our selection of data sets was based on the following criteria: importance for the discipline, size and region of its party population, degree of party information harmonization, and accessibility. Among these data sets are the main expert surveys (Bakker et al., 2015 (CHES); Benoit and Laver, 2006 (PPMD); Castles and Mair, 1984; Huber and Inglehart, 1995; Kitschelt, 2013; Morgan, 1976; Ray, 1999; Rohrschneider and Whitefield, 2012 (KUREP); Szöcsik and Zuber, 2015 (EPAC); Wiesehomeier and Benoit, 2009 (PPLA)), some of the major surveys (Afrobarometer, 2016; EES, 2014; ESS, 2002–2014; WVS, 1990–2014), well-established classifications of political parties (Coppedge, 1997; Janda, 1980; Thomas, 1975: 75), information about electoral performance (Cruz et al., 2016 (DPI); Mackie and Rose, 1991), and government participation (Andersson et al., 2014 (ERDDA)).

Creating a new party database based on existing data sources offers a fresh approach for a modern infrastructure on political parties. In creating such a database, we have faced some technical challenges and have developed new approaches to gathering data in the social sciences, which we present now.

Technical implementation and party-linking workflow

Three elements form the technical foundation of Party Facts: a database, a website, and a version control system. We extract party information from external data sets with software scripts and archive this information in a version control system. This data are imported into a relational database where information about core and external parties is stored, as well as some additional information on countries and data sets. A website provides an accessible interface to the information stored in the Party Facts database. It allows registered users to semi-automatically link political parties from external data sets to our new list of core parties and to edit the core parties. We present each of these technical elements in Figure 3 and in detail below.

Table 2. Data sets linked in Party Facts.

Name	Description	Years	Parties	Countries	Core parties
CLEA	Election results at the constituency level	1788–2017	2257	156	824
DPI	Database of political institutions	1975–2015	1678	171	179
PolCon	Political constraint index	1800–2012	1655	181	401
ParlGov	Parties, election results, and cabinets (EU and OECD)	1900–2018	1330	39	1113
Manifesto Project	Election program analysis for most democratic elections since 1945	1920–2017	1146	60	377
Mackie/Rose (1991)	Election results in Western nations	1800–1991	734	25	224
EJPR PDY	Election, referenda, governments, institutional reforms since the 1990s	1987–2015	655	37	15
WVS	Public opinion surveys on changing values	1990–2014	571	84	62
ESS	Public opinion surveys in Europe	2002–2014	546	31	19
Kitschelt (2013)	Expert survey on leadership accountability	2008	506	88	161
ERDDA (2013)	Cabinets, Parliaments, and Parties in postwar Europe	1944–2013	458	29	14
CHES	CHES on European integration, ideology, and policies	1999–2014	449	38	33
PPMD (2006)	Expert survey on party policy in modern democracies	2000–2004	370	47	8
Huber/Inglehart (1995)	Expert survey on parties' left/right positions	1990–1995	300	42	23
EPAC (2016)	Expert survey on ethno-nationalism in party competition	2011–2016	293	22	45
Coppedge (1997)	Classification of Latin American political parties	1912–1995	252	11	77
KUREP (2012)	Expert surveys in West and Central-Eastern Europe	2007	210	27	4
EES 2014	European Parliament Election Study 2014	2014	199	28	0
Jennings/Wlezien	Election polling information since the 1940s	1935–2017	198	31	4
Ray (1999)	Expert survey on European integration	1984–1996	191	17	0
PPLA (2009)	Expert survey parties and presidents in Latin America	2006	163	18	135
Janda (1980)	Political parties around the world—1950 to 1962	1950–1962	158	53	60
Morgan (1976)	Expert survey to study coalition formation	1919–1975	136	12	16
PPDB (2017)	Political Party Database Project	2002–2014	122	19	0
Afrobarometer (2016)	Public opinion surveys in Africa	2015	116	35	10
Thomas (1975)	Policy orientations of Western political parties	1870–1960	54	12	3

WVS: World Values Survey; CHES: Chapel Hill Expert Surveys; ESS: European Social Survey.
Note: Only the party information (names, size, years) from the respective data set is included into Party Facts. "Parties" shows the number of external parties imported from a data set and "core parties" the approximated number of new core parties derived from the data sets.

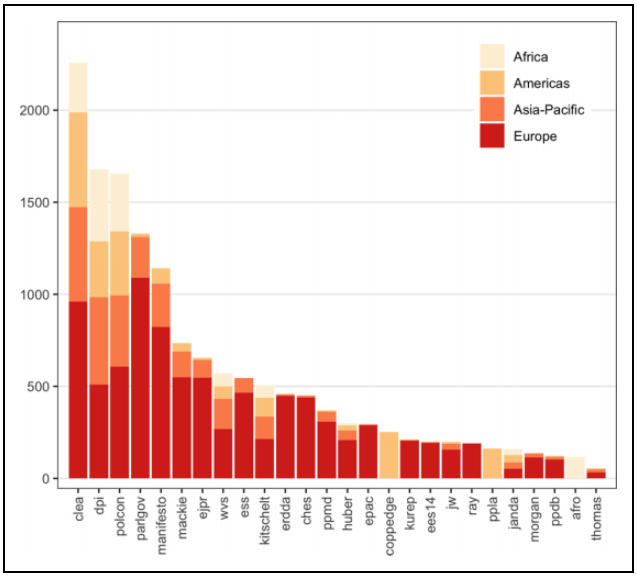

Figure 2. Number of parties from data sets. *Note*: Number of external parties in Party Facts by data set and world regions.

As a first step, we extract party information from an external data set with a software script. We do not include the entire original data set into Party Facts but extract only the required party information. This includes all information about a party name and ideally information about the party size and the period of its existence. In some data sets, this information is provided in an appendix; in others, it is included in the full data set from which we extract the respective columns. For example, the WVS includes an appendix about party codes and the respective party names. We combine this information with an approximation of the maximum party size that we calculate based on the number of respondents that choose the respective party in the vote intention question. For every external data set, an import script creates a new data set that includes only the party-level information (name, size, duration) that is imported into the database. Depending on the structure of the party information in the external data sets, we may have to manually preprocess the source data before the import. In fact, cleaning-up and harmonizing party information from data sets took a significant amount of time. During the harmonization of our party data, Wikipedia was an invaluable

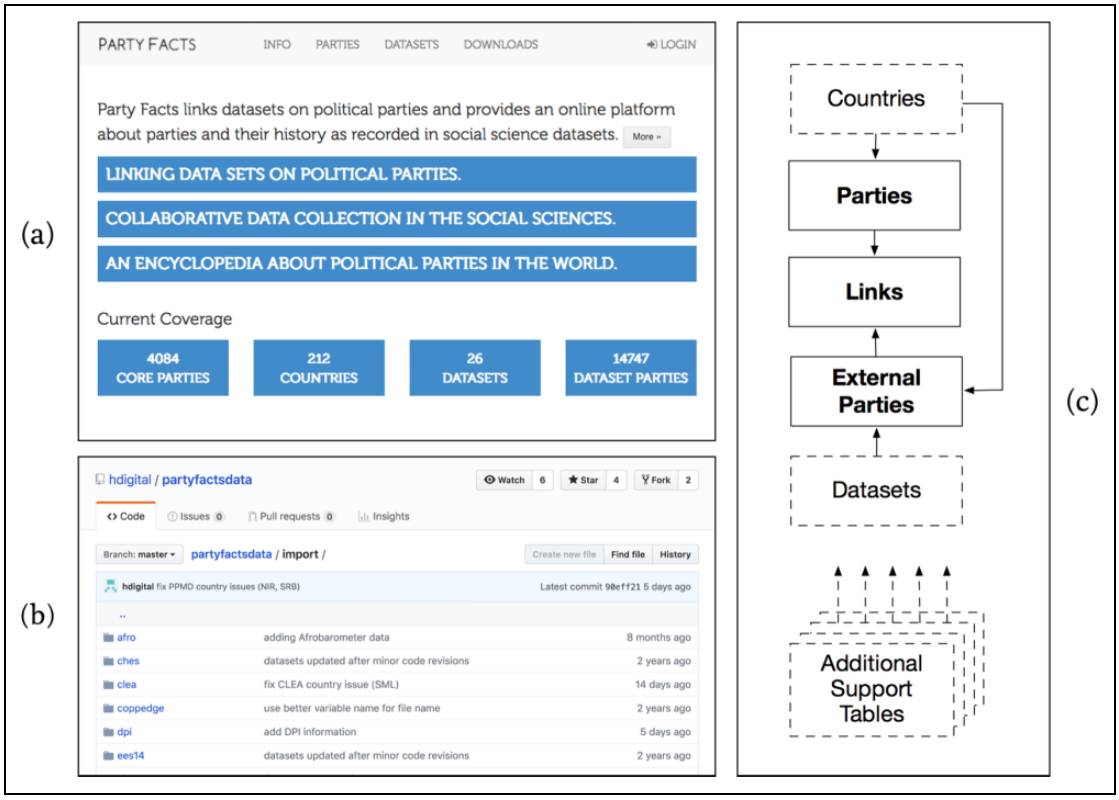

Figure 3. Technical foundations of Party Facts. *Note*: The three technical foundations of Party Facts: (a) the main page of the Party Facts website, (b) the version control system of the data set imports, and (c) an abstract scheme of the Party Facts database structure.

source for information about political parties across the globe.

We provide the option to archive the party-level information from the external source and the script we used to extract the party information in a version control system. Version control systems allow software developers to collaborate by sharing software code and have become increasingly popular in data science. They allow users to combine and archive different versions of code and data and to document the respective changes. In Party Facts, a version control system allows us to archive different versions of the party data we import and to cooperate with other social scientists.

The party-level information from an external data set is imported into the database where it can be linked with the parties in our core list. The database has three main tables and some additional support tables. One of the main tables includes all the core parties and another all parties imported from external data sets. In the latter, we include all original information, especially the unchanged variable names, in one entry and map the party-level information (name, size, duration) to the respective data columns. A third table is used to link the core parties and external parties. Some additional tables in the database provide information about the imported data sets, the countries in Party Facts, and other documentation. Most of these database details are hidden from users, since they access the database through the website or work with the merge tables we provide and archive.

On the website, we provide information about the core and external parties and about the links between them. The website is also the interface through which users can link parties and provides some semi-automatic approaches to make the linking process easier. We try to link external parties during the import into the database, although we do so only if there is one exact name match (abbreviation, original, or English name) between the imported external party and one core party. All other parties are linked on the website. We simplify this process by providing suggestions for these links based on record linkage. With a record-linkage algorithm, we determine and suggest the most likely match of an external party and a core party based on the textual similarity of names. This suggested link has to be confirmed by a user on the website before the respective link is added to the database. It is significantly easier and faster to link external parties through the record-linkage approach, and a high share of correct links is identified and suggested by this approach. Only the remaining parties need to be linked manually to the respective core party of the country. Commonly, this last step includes adding a new core party if the respective party does not yet exist in the Party Facts database.

The Party Facts website is very interactive. It shows which external parties are linked to a core party. For each external party, it also shows which other external parties it can be linked to through a core party. Users can add, edit, and correct information about core parties on the website, but the external parties cannot be edited on the website. These clean-ups and corrections of external parties are confined to the import script. The website provides support for linking parties from external data sets more easily. We have also drawn inspiration from collaborative online platforms such as Wikipedia. All user changes are logged and archived. An activity stream gives information about recent changes and users get credit for linking political parties. We hope that these features will encourage collaboration among social scientists on the Party Facts platform.

We provide two main tables for download on the Party Facts page and in a data archive. A merge table allows researchers to combine the external data sets. This table includes each external party from the database and the core party id it has been linked to. At this stage, the merge table already includes party ids for more than 25 data sets that we have harmonized for the initial version. A second table includes the list of core parties and the additional information for each core party that is available and editable on the Party Facts page. Time stamps for the initial creation of each observation and for the last modification document all changes in these two tables. This information and further documentation are also stored in a major long-term data archive (https://dataverse.harvard.edu/dataverse/partyfacts). These archived versions of the Party Facts data provide access independently of the main website.

How can others add data sets to the Party Facts project? It is our goal to establish a platform to link party information from existing data sources and to develop a more coherent record of political parties in the world. Hence, we rely on contributions from other social scientists to add to Party Facts. The technical stack we have presented here allows others to add party information from a data set through the version control system. There are already many existing examples that demonstrate how the party information can be extracted from a data set. The party information is added to the Party Facts database and the new party information imported can be linked with existing core parties through the interactive website.

Combining a database, a website and a version control system is highly beneficial. The database allows us to store the party information coherently. The website makes the information in the database accessible where it can be linked and modified by the users. The version control system stores and archives all relevant information documenting the changes between different versions. A long-term archive provides updated stable versions of the main data at regular intervals.

Cross-validating left-right positions

We conclude with a brief application of Party Facts for empirical studies of political representation. Today, most work on representation relies on different positional measures to determine the congruence of voters and parties. Powell (2009: 1476–80) discusses different approaches to assess left-right positions of citizens and parties. According to him, congruence can be measured by comparing public opinion surveys and expert surveys, on the basis of left-right positions and electoral results from the Manifesto data, or by comparing citizens' self-identification on a left-right scale with their perception of party positions. Increasingly, these approaches are combined. Ezrow et al. (2011: 280), for example, compare Manifesto left-right positions of political parties with Eurobarometer self-placements of voters to assess party positional moves.

For these studies on political congruence, linking party-level information from different sources has been a mundane and time-consuming endeavor. Resources needed to link potentially available party-level information have also limited the set of countries studied mainly to West European democracies. In addition, scholars have often combined party information for one particular set of studies only and the party data that have already been linked cannot easily be extended. It is exactly this set of regular challenges in applied work on political parties that Party Facts aims to address.

Here, we demonstrate how different measures of parties' left-right positions can be combined with Party Facts. We will not provide a full congruence study in line with Ezrow et al. (2011) but limit our application to demonstrate the validity of left-right measures across different data sets. We follow a format that was used earlier to validate expert surveys by comparing measures from different data sets (Bakker et al., 2015: 149).

We focus exemplary on a period in the late 2000s and choose two prominent public opinion surveys (ESS and WVS including EVS observations for WVS parties), two major expert surveys (CHES and Kitschelt), and the Manifesto positions. Two of the data sets have a broad global coverage (WVS and Kitschelt), two focus mainly on European countries (ESS and CHES), and the Manifesto data set is somewhere in between. For all data sets with multiple waves (ESS, WVS, CHES), we select the first wave between 2006 and 2010, and for the Manifesto data, we select the first election between 2006 and 2010. From this information, we extract the respective left-right measures. For public opinion surveys, we construct voters' party positions by combining the vote intention question and voters' self-placement on a left-right scale. This leads to voters' left-right positions for all parties included in the respective public opinion survey. For expert surveys, we select the respective left-right scale and we use the left-right measure

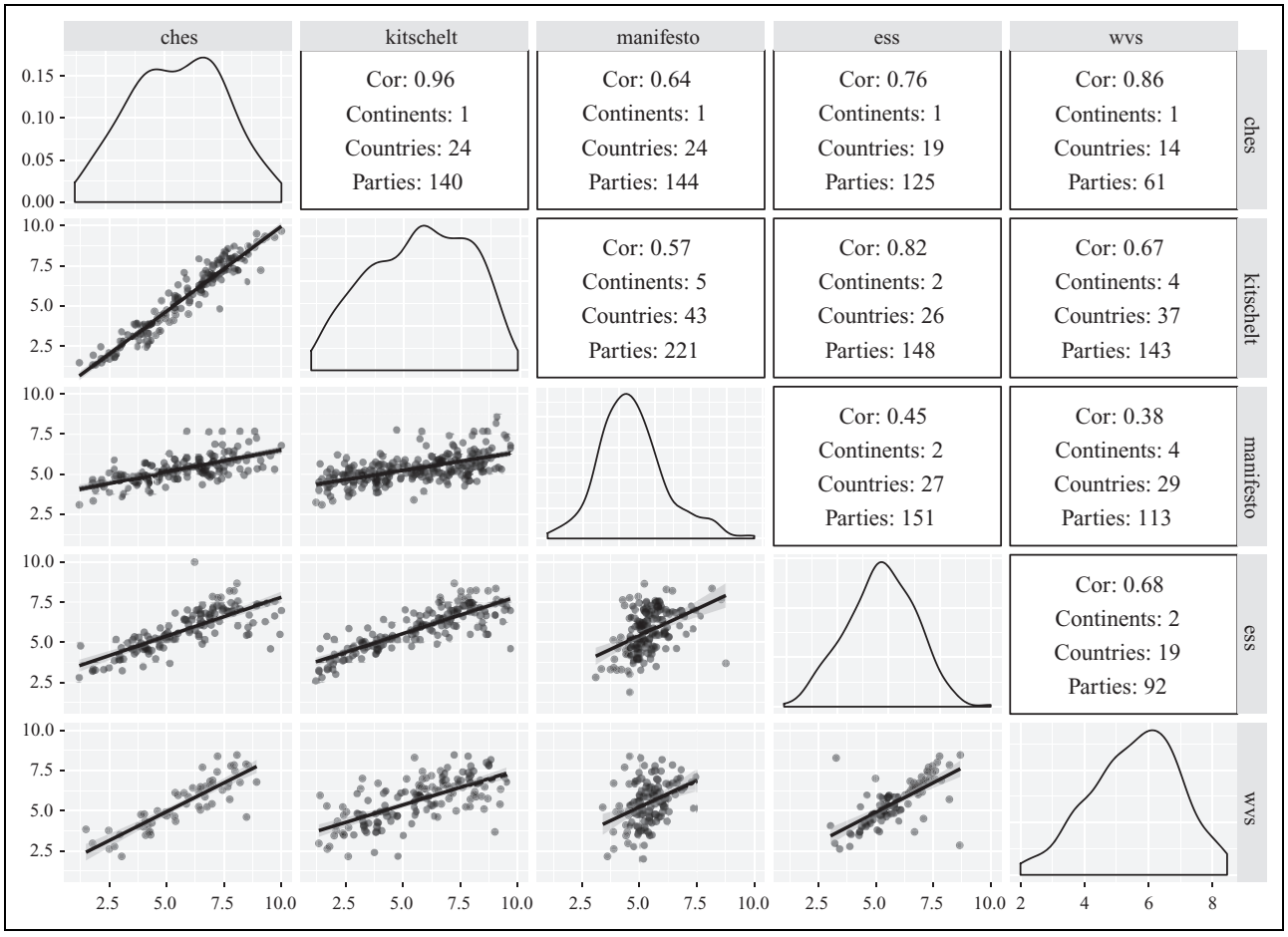

Figure 4. Example of cross-validating left-right positions with Party Facts. *Sources*: CHES (Bakker et al., 2015), Democratic Accountability and Linkages Project (Kitschelt, 2013), Manifesto Project (Volkens et al., 2018), ESS (European Social Survey, 2017), WVS (World Values Survey Association, 2015) including parts of EVS (EVS, 2015). CHES: Chapel Hill Expert Survey; ESS: European Social Survey; EVS: European Values Study; WVS: World Values Survey.

("rile") provided by the Manifesto Project. We rescale all five measures to fit a 1–10 scale.

Party Facts allows us to combine these different sources significantly more easily. In fact, we can link the more than 500 parties from these five data sets with little effort, a task that was previously time-consuming, mundane, and error prone. Figure 4 presents a cross-validation of the two public opinion surveys, two expert surveys, and the Manifesto data. In the lower left triangle, an individual scatter plot shows the left-right positions for each party that is included in both of two data sets. We show the correlation between the different data sources (summary scores in the upper right triangle and regression lines in the scatter plots in the lower left triangle) and their distribution in a density plot (diagonal) and present a summary of the number of parties, countries, and continents (upper right triangle) that are included in each respective comparison. The graph demonstrates that the different measures correspond well. We find a strong correlation between left-right measures in expert surveys and public opinion surveys in European countries. There is a weaker correlation of these measures with the "rile" from the Manifesto data, which could have various explanations. In contrast to the other approaches, Manifesto data measures self-ascribed party positions. The Manifesto "rile" measure is not context specific to elections or countries but aims at providing scores that are directly comparable across time and space. There is a long debate with multiple other proposals for extracting left-right positions from the Manifesto data. Using a more context-specific operationalization of left-right positions based on the Manifesto data would result in a higher congruence (Franzmann, 2015). In addition, the density plot informs us that voters place themselves more centrally than political parties are placed by experts. Finally, we see that the number of countries the respective data sets cover varies significantly. Our results are in line with existing results that compare party positions for European democracies (Bakker et al., 2015: 150).

Comparing the left-right positions from different sources is only one of the many potential applications for

Party Facts to combine party-related information. We could also use party-level information about the government/opposition status of parties (ERDDA, EJPR-PDY, ParlGov), district-level election results (CLEA), organizational power structures (Janda), or membership information (PPDB). New party information from other data sets can easily be added through the approach we have developed. Finally, the information we combine provides the basis for a comprehensive list of all relevant political parties in the world as recorded in social science data sets.

Conclusion

A project like Party Facts raises crucial questions about the sustainability of such an endeavor that relies heavily on modern web technologies. We aim to further enhance and extend the project. The data that are currently included and harmonized, as well as the linked data sets we discuss in this article, are permanently archived for long-term storage. The website that displays the project information, presents the data, allows user-based linking of new data sets, enables link validation by users, and supports downloading the data is publicly accessible. It has been used for several years and we plan to maintain it, although the website may be replaced in the future with a different user interface. It is important to distinguish the harmonized and linked party data that we provide, the new approach to data collection we present in this article, and the prototype website we have developed and used over the last few years.

For the future, we have two potential extensions of the project in mind that would significantly increase the value of an online database on political parties. First, we would like to specify information about the population of political parties in Party Facts by adding information about the links between parties and their naming history. Conceptually, it is important to distinguish entirely new parties from those that are formed on the basis of established parties. This information is traditionally available as descriptive information about political parties and is provided in handbooks and almanacs. A comprehensive digital source about the population of political parties would profit from this information as well. Information about the origin of a party is also important to develop a more coherent definition and operationalization of political parties in order to systematically distinguish old and new parties.

Second, we would like to add a user survey for each party to Party Facts. Some information about political parties cannot be established through coherent coding criteria. However, particularly for new parties, some information about their positions and issue saliences is of relevance for applied work in the social sciences. In addition, information about potential party families and some classifications, such as populism or anti-system parties, could be collected through a user survey. Such a survey could be added to each party and the online interface would allow users to answer a few key questions about that particular party. Adding these two features to Party Facts in the future would require input from experts on political parties and a more in-depth discussion of new coding rules.

The success of the project will mainly depend on finding a critical mass of contributors to the project. Recent innovations in online technologies have made such an approach feasible. Nevertheless, such a project still requires a number of contributors who add regularly to the project, and a large number of country experts who can provide crucial information about political parties. As of today, the thousands of parties from more than 25 external data sets in 212 countries and the merge tables we provide are already a valuable new source that allows social scientists to tackle new research questions.

Acknowledgements

For comments, fruitful discussions, and encouragement, we would like to thank Pola Lehmann, Nicolas Merz, Theres Matthieß (all WZB Berlin), Tomasz Żóltak (IBE Warsaw), and Christina Zuber (BIGSSS Bremen). Holger Döring benefited from a research stay at the Center for Political Studies, University of Michigan in 2015, and thanks Allen Hicken, Ken Kollman, Yioryos Nardis (University of Michigan), Kevin Deegan-Krause (Wayne State), and Thomas Mustillo (Notre Dame). Paul Bederke, Jan Fischer, Phillip Hocks, and Jan Schwalbach provided crucial research assistance at the University of Bremen. Ingo Rohlfing and Nils Düpont provided resources to extend the coverage of the database. Three anonymous reviews have helped us to present the content in a significantly more accessible way.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the University of Bremen through the M8 Initiative.

ORCID iD

Holger Döring https://orcid.org/0000-0002-6616-8805

References

Alexander R (1982) *Political Parties of the Americas: Canada, Latin America, and the West Indies*. Westport: Greenwood Press.

Andersson S, Bergman T and Ersson S (2014) *The European representative democracy data archive, Release 3*. Stockholm: Main sponsor: Riksbankens Jubileumsfond (In2007-0149: 1-E). Available at: www.erdda.se.

Bakker R, de Vries C, Edwards E, et al. (2015) Measuring party positions in Europe: the Chapel Hill expert survey trend file, 1999–2010. *Party Politics* 21(1): 143–152.

Barnea S and Rahat G (2011) Out with the old, in with the 'new': What constitutes a new party? *Party Politics* 17(3): 303–320.

Benoit K and Laver M (2006) *Party Policy in Modern Democracies*. London, New York: Routledge.

Budge I, Klingemann H-D, Volkens A, et al. (2001) *Mapping Policy Preferences: Estimates for Parties, Electors, and Governments, 1945–1998*. Oxford: Oxford University Press.

Bugajski J (2002) *Political Parties of Eastern Europe: A Guide to Politics in the Post-Communist Era*. Armonk: M.E.Sharpe.

Castles FG and Mair P (1984) Left right political scales: some expert judgments. *European Journal of Political Research* 12(1): 73–88.

Coppedge M (1997) A classification of Latin American political parties. Kellogg Institute for International Studies – Working Paper #244. Available at: https://kellogg.nd.edu/documents/1539 (accessed 3 December 2015).

Cruz C, Keefer P and Scartascini C (2016) The Database of Political Institutions 2015 (DPI2015). Available at: https://publications.iadb.org/handle/11319/7408 (accessed 7 November 2018).

Döring H (2013) The collective action of data collection: a data infrastructure on parties, elections and cabinets. *European Union Politics* 14(1): 161–178.

Döring H (2016) Mapping established democracies: integrated data on parties, elections and cabinets. *Electoral Studies* 44: 535–543.

Döring H and Manow P (2018) Parliaments and governments database (ParlGov): Stable version 2018 – 01 May 2018. DOI: 10.7910/DVN/F0YGNC Harvard Dataverse, V1, UNF:6:NQUZONiGgflT9PBAyCVu+Q== [fileUNF].

European Social Survey (2017) ESS cumulative file rounds 1-7, ed.1.0 – NSD - Norwegian Centre for Research Data, Norway – Data archive and distributor of ESS data for ESS ERIC. Available at: www.europeansocialsurvey.org (accessed 13 September 2017).

EVS (2015) European Values Study Longitudinal Data File 1981-2008 (EVS 1981-2008). GESIS Data Archive, Cologne. ZA4804 Data File Version 3.0.0. DOI: 10.4232/1.12253.

Ezrow L, Vries CD, Steenbergen M, et al. (2011) Mean voter representation and partisan constituency representation: Do parties respond to the mean voter position or to their supporters? *Party Politics* 17(3): 275–301.

Franzmann S (2015) Towards a real comparison of left-right indices: a comment on Jahn. *Party Politics* 21(5): 821–828.

Henisz WJ (2000) The institutional environment for economic growth. *Economics and Politics* 12(1): 1–31.

Huber JD and Inglehart R (1995) Expert interpretations of party space and party locations in 42 societies. *Party Politics* 1(1): 73–111.

Jacobs F (ed.) (1989) *Western European Political Parties: A Comprehensive Guide*. Harlow: Longman.

Janda K (1980) *Political Parties: A Cross-National Survey*. New York: The Free Press.

Katz R and Mair P (1992) *Party Organizations: A Data Handbook on Party Organizations in Western Democracies, 1960-90*. London: Sage.

Kitschelt H (2013) *Dataset of the Democratic Accountability and Linkages Project (DALP)*. Raleigh: Duke University. Available at: http://sites.duke.edu/democracylinkage (accessed 12 June 2014).

Klingemann HD, Volkens A, Bara J, et al. (2006) *Mapping Policy Preferences II: Estimates for Parties, Electors and Governments in Central and Eastern Europe, European Union and OECD 1990-2003*. Oxford: Oxford University Press.

Kollman K, Hicken A, Caramani D, et al. (2018) Constituency-Level Elections Archive (CLEA). 2018 Version (20180507). Ann Arbor: Center for Political Studies, University of Michigan [producer and distributor]. Available at: http://www.electiondataarchive.org/ (accessed 23 May 2018).

Laver M and Hunt WB (1992) *Policy and Party Competition*. New York: Routledge.

Mackie T and Rose R (1991) *The International Almanac of Electoral History*. Washington: Congressional Quarterly.

Morgan MJ (1976) *The Modelling of Governmental Coalition Formation: A Policy-based Approach with Interval Measurement*. Ann Arbor: University of Michigan.

Mustillo T and Springer JA (2014) Seeing beyond the datasets: the unexploited structure of electoral and institutional data. *International Journal of Social Research Methodology* 18(4): 349–365.

Nohlen D and Stöver P (eds.) (2010) *Elections in Europe: A Data Handbook*. Baden-Baden: Nomos Verlagsgesellschaft.

Poguntke T, Scarrow SE, Webb PD, et al. (2016) Party rules, party resources and the politics of parliamentary democracies: How parties organize in the 21st century. *Party Politics* 22(6): 661–678.

Powell GB Jr (2009) The ideological congruence controversy: the impact of alternative measures, data, and time periods on the effects of election rules. *Comparative Political Studies* 42(12): 1475–1497.

Ray L (1999) Measuring party orientations towards European integration: results from an expert survey. *European Journal of Political Research* 36(2): 283–306.

Rohrschneider R and Whitefield S (2012) *The Strain of Representation: How Parties Represent Diverse Voters in Western and Eastern Europe*. Oxford: Oxford University Press.

Rose R and Munro N (2003) *Elections and Parties in New European Democracies*. Washington: CQ Press, A Division of Congressional Quarterly.

Sager D (2009) *Political Parties of the World*, 7th ed. London: John Harper.

Schedler A (2012) The measurer's dilemma: coordination failures in cross-national political data collection. *Comparative Political Studies* 45(2): 237–266.

Stöss R (1986) *Parteien-Handbuch: Die Parteien Der Bundesrepublik Deutschland, 1945-1980. Sonderausgabe*. Opladen: Westdeutscher Verlag.

Szöcsik E and Zuber C (2015) EPAC: a new dataset on ethnonationalism in party competition in 22 European democracies. *Party Politics* 21(1): 153–160.

Thomas JC (1975) *The Decline of Ideology in Western political parties: A study of changing policy orientations*. Beverly Hills, CA: Sage Professional Papers in Contemporary Political Sociology.

Volkens A, Bara J, Budge I, et al. (2013) *Mapping Policy Preferences from Texts III: Statistical Solutions for Manifesto Analysts*. Oxford: Oxford University Press.

Volkens A, Lehmann P, Matthieß T, et al. (2018) *The manifesto data collection. Manifesto Project (MRG/CMP/MARPOR). Version 2018a*. Berlin: Wissenschaftszentrum Berlin für Sozialforschung (WZB). DOI: 10.25522/manifesto.mpds.2018a.

van de Wardt M and Otjes S (2018) *Irrational parties, rational voters? Examining how new parties and voters react to party voter incongruence*. Paper presented at the ECPR General Conference, Hamburg, 22–25 August 2018.

Wiesehomeier N and Benoit K (2009) Presidents, parties and policy competition. *The Journal of Politics* 71(4): 1435–1447.

World Values Survey Association (2015) *WVS 1981-2014 longitudinal aggregate V.20150418*. Madrid: Aggregate file producer. JD Systems. Available at:www.worldvaluessurvey.org. (accessed 13 September 2017).

Author biographies

Holger Döring is a senior researcher in political science and a member of the SOCIUM Research Center at the University of Bremen. His research focuses on political institutions and democratic representation in advanced democracies and the European Union. He is the main author of the ParlGov data infrastructure on parties, elections, and cabinets (http://www.parlgov.org).

Sven Regel is a PhD researcher at the WZB Berlin Social Science Center. He works on parliamentary behavior, political representation, party competition, and quantitative text analysis. He is a core member of the Manifesto Project research group (https://manifesto-project.wzb.eu).